Explanation of Sales Prediction Script

Created by Jangita Takeshwar

This script is designed to analyze the impact of advertising on sales using Linear Regression. It utilizes a dataset containing advertising expenses on TV, radio, and newspapers to predict sales performance.

1. Import Libraries

The script starts by importing necessary Python libraries:
- pandas and numpy for data manipulation
- matplotlib and seaborn for data visualization
- scikit-learn modules for model training and evaluation.

2. Load Dataset

The dataset is read from a CSV file using pandas. It contains advertising expenditures and corresponding sales data.

3. Display Data and Check for Missing Values

The script displays the first few rows of the dataset and checks for any missing values using the `isnull().sum()` function.

4. Define Features and Target Variable

The independent variables (features) include TV, Radio, and Newspaper expenses. The dependent variable (target) is 'Sales'.

5. Split Data into Training and Testing Sets

The dataset is split into training (80%) and testing (20%) sets using `train_test_split`. This ensures the model is trained on one portion and evaluated on another to measure performance.

6. Train the Linear Regression Model

A Linear Regression model is initialized and trained using the training dataset. It learns the relationship between the advertising expenses and sales.

7. Make Predictions

The trained model predicts sales based on the test dataset. The predicted values are stored in `y_pred`.

8. Evaluate the Model

The model's performance is assessed using evaluation metrics:
- Mean Absolute Error (MAE)

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score (coefficient of determination).

9. Visualize Predictions vs Actual Sales

A scatter plot is created to compare actual vs predicted sales. This helps assess the accuracy of the model visually.